

# A Note on the Malleability of the El Gamal Cryptosystem

Douglas Wikström

Swedish Institute of Computer Science (SICS)  
douglas@sics.se

**Abstract.** The homomorphic property of the El Gamal cryptosystem is useful in the construction of efficient protocols. It is believed that only a small class of transformations of cryptotexts are feasible to compute. In the program of showing that these are the only computable transformations we rule out a large set of natural transformations.

## 1 Introduction

Several efficient cryptographic protocols are based on the El Gamal cryptosystem. The reasons for this are mainly the algebraic simplicity of the idea, and the homomorphic property it possesses. The latter property makes the El Gamal system malleable, i.e. given  $c = E(m)$  it is feasible to compute  $c' = E(f(m))$ , for some nontrivial function  $f$ .

It is commonly conjectured that the El Gamal cryptosystem is malleable only for a small class of simple functions, and this is sometimes used implicitly in arguments about the security of protocols. Thus it is an important problem to characterize the malleability of the El Gamal cryptosystem. We take a first step in this direction.

We formalize the problem, and discuss why restrictions of the problem are necessary. Then we show that the only transformations that can be computed perfectly are those of a well known particularly simple type. Further on we give two examples that show that possible future results are not as strong as we may think. Finally we rule out a large set of natural transformations from being computable.

### 1.1 The El Gamal Cryptosystem

First we review the El Gamal cryptosystem and introduce some notation.

All computations of the El Gamal cryptosystem [2] take place in a group  $G$ , such as a subgroup of  $\mathbb{Z}_p^*$  or an elliptic curve group. We assume that  $|G| = q$  is prime, and that there is a system wide generator  $g$  of  $G$ .

Keys are generated by choosing  $x \in \mathbb{Z}_q$  uniformly and computing  $y = g^x$ , where the private key is  $x$  and the public key is  $y$ . Encryption is defined by  $E_y(m, r) = (g^r, y^r m)$  for a message  $m \in G$  and a random exponent  $r \in \mathbb{Z}_q$  chosen uniformly. Decryption is defined by  $D_x(u, v) = vu^{-x}$ .

Above we described the system generically for any group of prime order, but to define security we need to consider families of groups. Let  $q_1, q_2, \dots$  be an increasing sequence of prime numbers, where  $\lceil \log_2 q_n \rceil = n$ , and let  $q = \{q_n\}$ . We denote the family  $\{\mathbb{Z}_{q_n}\}$  by  $\mathbb{Z}_q$ . Let  $G = \{G_n\}$  be a family of groups such that  $|G_n| = q_n$ . We use  $\mathbb{Z}_q$  and  $G$  generically to denote  $\mathbb{Z}_{q_n}$  and  $G_n$  when no reference to  $n$  is made. We take  $n$  to be the security parameter.

Finally we assume that the Decision Diffie-Hellman assumption (DDH) [1, 4] holds in  $G = \{G_n\}$ . Let  $g$  be a generator in  $G$ , and let  $a, b$  and  $c$  be uniformly and independently distributed in  $\mathbb{Z}_q$ . Then DDH states that  $(g^a, g^b, g^{ab})$  and  $(g^a, g^b, g^c)$  are indistinguishable. Tsionis and Yung [4] formally proves that this implies that the El Gamal cryptosystem is semantically secure over  $G$ .

### 1.2 Notation

Throughout we denote by PC the set of polynomial size circuit families. We abbreviate uniformly and independently distributed by u.i.d. .

To simplify notation we avoid using the security parameter  $n$ . For example, if we consider a family of functions  $\phi = \{\phi_n\}$ , we use  $\phi$  generically for  $\phi_n$  and each appropriate  $n$ .

### 1.3 The Problem

For any message  $m \in G$  there exists a unique element  $m_e \in \mathbb{Z}_q$  such that  $m = g^{m_e}$ . Thus any El Gamal encryption  $(g^r, y^r m)$  can be written  $(g^r, y^r g^{m_e})$ . The latter notation, is sometimes more natural and we use both conventions.

There is a small and well know class of transformations of cryptotexts, used in many protocols, that we summarize in an observation.

**Observation 1.** *Set  $\phi(r) = b_1 r + b_0$  and  $\psi(m_e) = b_1 m_e + h_0$ . Then the map:*

$$(g^r, y^r g^{m_e}) \mapsto (g^{\phi(r)}, y^{\phi(r)} g^{\psi(m_e)})$$

*is feasible to compute.*

Some authors use the term homomorphic cryptosystem, since these transformations can be formulated as group homomorphisms.

It is natural to ask what other transformations can or can not be computed “under encryption”. For simplicity we use the non-uniform computational model, i.e. feasible transformations are transformations that can be computed by a deterministic non-uniform circuit family.

We restrict our attention to deterministic transformations, since given a probabilistic algorithm that computes a transformation there is a deterministic circuit family that performs at least as well.

Given  $y = g^x$ , each pair  $(u, v) \in G \times G$  can be uniquely represented on the form  $(u, v) = (g^r, y^r g^{m_e})$ . This implies that for each function  $f : G \times G \rightarrow G \times G$ , and  $y \in G$ , there are unique functions  $\phi_y, \psi_y : \mathbb{Z}_q \times \mathbb{Z}_q \rightarrow \mathbb{Z}_q$ , such that:

$$f(u, v) = f(g^r, y^r g^{m_e}) = (g^{\phi_y(r, m_e)}, y^{\psi_y(r, m_e)} g^{\psi_y(r, m_e)}) .$$

Most general functions  $f$  are not what we intuitively would consider “transformations computed under encryption”, and it seems difficult to prove anything useful if we consider any function  $f$  a transformation of cryptotexts.

Our approach is therefore to require that a transformation is given by a fixed pair  $(\phi, \psi)$  of deterministic functions  $\phi, \psi : \mathbb{Z}_q \times \mathbb{Z}_q \rightarrow \mathbb{Z}_q$  and parametrized by  $y$ , i.e. we define a map  $(y, \phi, \psi) : G \times G \rightarrow G \times G$  for each  $y$  by the following:

$$(y, \phi, \psi) : (g^r, y^r g^{m_e}) \mapsto (g^{\phi(r, m_e)}, y^{\phi(r, m_e)} g^{\psi(r, m_e)}) .$$

Such transformations act uniformly for all  $y$ , i.e. given  $(u_i, v_i) = (g^r, y_i^r g^{m_e})$  for  $i = 1, 2$  we have  $(y_i, \phi, \psi)(u_i, v_i) = (g^{\phi(r, m_e)}, y_i^{\phi(r, m_e)} g^{\psi(r, m_e)})$ .

Our method can not be applied to general uniform transformations, and we are forced to further restrict the problem. We require that  $\phi$  depends only on  $r$ , and that  $\psi$  depends only on  $m_e$ . Thus we study the special problem posed as follows:

*Problem 1.* Given  $\phi, \psi : \mathbb{Z}_q \rightarrow \mathbb{Z}_q$ , let  $(y, \phi, \psi)(g^r, y^r g^{m_e}) = (g^{\phi(r)}, y^{\phi(r)} g^{\psi(m_e)})$ . For which  $\phi$  and  $\psi$  is the transformation  $(y, \phi, \psi)$  feasible to compute?

## 2 Our Results

We exhibit two propositions. The first shows that only transformations of the type described in Observation 1 can be computed perfectly. Then we give two examples that show that strong results can not be hoped for. Finally we give the main proposition, which may have some practical significance. It identifies a set of functions  $\psi$  such that the map  $(y, \phi, \psi)$  is hard to compute for every  $\phi$ .

### 2.1 Some Preparation

The hypothesis of the propositions differ only slightly depending on which case is considered. To avoid duplication of the hypothesis, and for increased clarity we give it here.

#### Hypothesis 1.

1. Let  $G = \{G_n\}$  be a family of groups such that  $|G_n| = q_n$ , where  $q_n$  is a prime number such that  $\lceil \log_2 q_n \rceil = n$ , and assume that DDH holds in  $G$ . Let  $g = \{g_n\}$  be a generator of  $G$ .
2. Let  $X = \{X_n\}$  be a family of random variables, where  $X_n$  is u.i.d. in  $\mathbb{Z}_{q_n}$ , and let  $Y = \{Y_n\}$ , where  $Y_n = g^{X_n}$ .
3. Let  $R = \{R_n\}$  be a family of random variables, where  $R_n$  is u.i.d. in  $\mathbb{Z}_{q_n}$ .
4. Let  $M = \{M_n\}$  be a family of random variables on  $G_n$ , and define the induced family  $(U, V) = \{(U_n, V_n)\}$  of random variables by setting  $(U_n, V_n) = E_{Y_n}(M_n, R_n)$ .

5. Let  $\phi = \{\phi_n\}$  and  $\psi = \{\psi_n\}$  be families of functions over  $\mathbb{Z}_q$ , i.e.  $\phi_n, \psi_n : \mathbb{Z}_{q_n} \rightarrow \mathbb{Z}_{q_n}$ . Define for each family  $y = \{y_n\} \in G$  a family of maps  $(y, \phi, \psi) = \{(y_n, \phi_n, \psi_n)\}$ , where:

$$\begin{aligned} (y_n, \phi_n, \psi_n) : G_n \times G_n &\rightarrow G_n \times G_n \\ (y_n, \phi_n, \psi_n) : (g_n^r, y_n^r g_n^{m\epsilon}) &\mapsto (g_n^{\phi_n(r)}, y_n^{\phi_n(r)} g_n^{\psi_n(m\epsilon)}) . \end{aligned}$$

Definitions of  $M$ ,  $\phi$ , and  $\psi$  are given separately in each proposition. The following definition, first given by Goldwasser and Micali [3] define what should be considered randomly guessing the output of a knowledge function.

**Definition 1.** Let  $M = \{M_n\}$  be a family of random variables, where the outcomes of  $M_n$  are in  $G_n$ , and let  $f = \{f_n\}$  be a family of functions  $f_n : G_n \rightarrow G_n$ . We define:

$$p_n(f, M) = \max_{v \in G_n} \Pr[M_n \in f_n^{-1}(v)] .$$

The probability  $p_n(f, M)$  is the maximum probability of any algorithm to guess  $f_n(M_n)$  using no information on the outcome of  $M_n$  except its distribution.

Since El Gamal is semantically secure [3, 4] we have under the assumptions in the hypothesis and with arbitrary  $f = \{f_n\}$ , that  $\forall A \in \text{PC}, \forall c, \exists n_0$  such that for  $n > n_0$  it holds that:

$$\Pr[A(Y, (U, V)) = f(M)] < p_n(f, M) + \frac{1}{n^c} .$$

## 2.2 The Perfect Case

The following proposition says that if we require a circuit family to succeed with probability 1 in computing the map  $(y, \phi, \psi)$  the only possible maps are those where  $\psi$  is linear.

**Proposition 1.** Let  $G, X, Y, M, (U, V), \phi$  and  $\psi$  be as in Hypothesis 1, let  $M$  be arbitrarily distributed in  $G$ , and assume that  $\psi_n(x)$  is non-linear for infinitely many  $n$ .

Then  $\forall A \in \text{PC}, \exists n_0$  such that  $\forall n > n_0$ :

$$\Pr[A(Y, (U, V)) = (Y, \phi, \psi)(U, V)] < 1 .$$

*Proof.* The proof is by contradiction. Assume that  $A, \phi$ , and  $\psi$  as above show the proposition false for indices  $n$  in some infinite index set  $\mathcal{N}$ . Then  $\psi_1(x) = \psi(1+x) - \psi(x)$  is not constant. Let  $g^{m_0}$  and  $g^{m_1}$  be two messages such that  $\psi_1(m_0) \neq \psi_1(m_1)$ . Let  $A'$  be the circuit family that given a public key  $y$  and an encryption  $(u, v)$  of the message  $g^{m_b}$  computes  $(u_0, v_0) = A(y, (u, v))$  and  $(u_0, v_1) = A(y, (u, vg))$ , and returns  $b$  when  $v_1/v_0 = g^{\psi_1(m_b)}$ .

Clearly  $A'$  breaks the polynomial indistinguishability, and thus the semantic security of the El Gamal cryptosystem.  $\square$

### 2.3 Two Examples of Possible Approximations

In general we expect that the difficulty of computing a map  $(y, \phi, \psi)$  depends on both  $\phi$  and  $\psi$ . On the other hand, in applications we are more interested in how an adversary can transform the cleartext hidden inside a cryptotext. In most situations we expect the adversary to rerandomize its output, but as explained in Section 1.3 such an adversary implies the existence of a deterministic adversary. Thus, given a fixed  $\psi$ , a reasonable goal is to bound the probability for any adversary to compute  $(y, \phi, \psi)$  for any choice of  $\phi$ .

We now present two examples that show that we should not hope for general strong results. Both examples assume that  $G, X, Y, M, (U, V), \phi$  and  $\psi$  are as in Hypothesis 1, and that  $M$  is u.i.d. .

*Example 1.* Let  $\psi$  be arbitrary but fixed and let  $w$  maximize  $\Pr[M \in \psi^{-1}(w)]$ . Let  $A$  be the circuit family that computes  $r' = h(u)$ , where  $h : G \rightarrow \mathbb{Z}_q$ , and then outputs  $(g^{r'}, y^{r'} g^w)$ .

Clearly  $\Pr[A(Y, (U, V)) = (Y, \phi, \psi)(U, V)] = p_n(\psi, M)$ , where  $\phi(r) = h(g^r)$ . The example shows that for every  $\psi$  there is a non-trivial  $\phi$  such that the map  $(y, \phi, \psi)$  can be computed with probability at least  $p_n(\psi, M)$ .

Thus the best general result under Hypothesis 1 we could hope for at this point is to show that  $\forall A \in \text{PC}, \forall c > 0, \exists n_0 > 0$ , such that for  $n > n_0$ :

$$\Pr[A(Y, (U, V)) = (Y, \phi, \psi)(U, V)] < p_n(\psi, M) + \frac{1}{n^c} ,$$

but no such result exists as the next example shows.

*Example 2.* Let  $c > 0$  be fixed and define  $B_n = \{x \in \mathbb{Z}_{q^n} : 0 \leq x \leq \frac{q^n}{n^c}\}$ ,  $B = \{B_n\}$ . Define  $\psi_n(x) = x + 1$  if  $x \in B_n$ , and  $\psi_n(x) = x^2$  otherwise, and set  $\phi = \text{id}$ . Let  $A$  be the circuit family that assumes that the input  $(u, v) = (g^r, y^r g^{m_e})$  satisfies  $m_e \in B$ , and simply outputs  $(u, vg)$ .

We have  $|\psi^{-1}(x)| \leq 3$  for all  $x \in \mathbb{Z}_q$ , which implies  $p_n(\psi, M) \leq \frac{3}{q^n}$ , but still  $A$  computes  $(y, \phi, \psi)$  with probability  $1/n^c$  for a fixed  $c$ . Thus the example shows that we can sometimes compute a transformation with much greater probability than  $p_n(\psi, M)$ , i.e. the probability of guessing  $\psi(m_e)$ .

Intuitively the problem seems to be that our ability to compute transformations from the class described in Observation 1 changes what should be considered guessing.

### 2.4 A Class of Hard $\psi$

We now exhibit a class of  $\psi$  that are hard in the sense that the map  $(y, \phi, \psi)$  is hard to compute for all  $\phi$ .

The idea of Proposition 2 below is that given input  $(y, (u, v))$  and an oracle  $A$  for computing a transformation  $(y, \phi, \psi)$  we can ask  $A$  several different but related questions. If  $A$  answers our questions correctly we are able to compute some derived knowledge function  $f$  of the cleartext.

Let  $\psi = \{\psi_n\}$  be a family of functions,  $\psi_n : \mathbb{Z}_{q_n} \rightarrow \mathbb{Z}_{q_n}$ , and let  $s \in \mathbb{Z}_q$ . Denote by  $\psi_s$  the function given by  $\psi_s(x) = \psi(x + s) - \psi(x)$ . We prove below that a  $\psi$  that satisfies the following definition has the desired property.

**Definition 2.** Let  $\psi = \{\psi_n\}$  be a family of functions,  $\psi_n : \mathbb{Z}_{q_n} \rightarrow \mathbb{Z}_{q_n}$ , let  $M = g^{M_e}$ , where  $M_e$  is a random variable in  $\mathbb{Z}_q$ , and let  $S$  be u.i.d. in  $\mathbb{Z}_q$ .

If  $\forall c > 0, \exists n_0 > 0$  such that  $\forall n > n_0$  we have:

$$\Pr[p_n(\psi_S, M) < \frac{1}{n^c}] > 1 - \frac{1}{n^c} ,$$

then we say that  $\psi$  is strongly non-linear with respect to  $M$ .

The following definition may seem more natural to some readers.

**Definition 3.** Let  $\psi = \{\psi_n\}$  be a family of functions,  $\psi_n : \mathbb{Z}_{q_n} \rightarrow \mathbb{Z}_{q_n}$ , let  $M_e$  and  $S$  be random variables in  $\mathbb{Z}_q$ , where  $S$  is u.i.d. .

If  $\forall a \in \mathbb{Z}_q, \forall c > 0, \exists n_0$  such that  $\forall n > n_0$  we have:

$$\Pr[\psi(M_e + S) - \psi(M_e) = \psi(S) + a] < \frac{1}{n^c} ,$$

then we say that  $\psi$  is strongly non-linear\* with respect to  $M_e$ .

Unfortunately it captures a larger class than Definition 2 as Lemma 1 below shows, and we can not prove Proposition 2 for all  $\psi$  satisfying this definition.

The essential difference between the two definitions is that in the second  $a$  is fixed, and does not depend on  $s$ , whereas in the first  $p_n(\psi_s, M)$  is maximized for each  $s$  independently. Note that if we fix  $S = s$  in the second definition there is always an  $a$  such that the resulting conditioned probability equals  $p_n(\psi_s, M)$ , but in general  $a$  depends on  $s$ .

**Lemma 1.** Strongly non-linear implies strongly non-linear\*.

*Proof.* Set  $J(S) = p_n(\psi_S, M)$ . Then  $\forall c > 0, \exists n_0$  such that  $\forall n > n_0$ :

$$\begin{aligned} & \Pr[\psi(M_e + S) - \psi(M_e) = \psi(S) + a] \\ &= \sum_{s \in \mathbb{Z}_q} \Pr[S = s] \Pr[\psi(M_e + s) - \psi(M_e) = \psi(s) + a] \\ &\leq \sum_{s \in \mathbb{Z}_q} \Pr[S = s] J(s) = \mathbb{E}[J(S)] \\ &= \Pr[J(S) < \frac{1}{n^c}] \mathbb{E}[J(S) | J(S) < \frac{1}{n^c}] + \Pr[J(S) \geq \frac{1}{n^c}] \mathbb{E}[J(S) | J(S) \geq \frac{1}{n^c}] \\ &< 1 \cdot \frac{1}{n^c} + \frac{1}{n^c} \cdot 1 = \frac{2}{n^c} . \end{aligned}$$

□

**The Main Proposition.** Informally the proposition below says that if  $\psi$  is strongly non-linear, then  $(y, \phi, \psi)$  is hard to compute for all  $\phi$ .

**Proposition 2.** Let  $G, X, Y, M, (U, V), \phi$  and  $\psi$  be as in Hypothesis 1, let  $M$  be u.i.d. in  $G$ , and assume that  $\psi$  is strongly non-linear with respect to  $M$ .

Then  $\forall A \in \text{PC}, \forall c > 0, \exists n_0 > 0$ , such that for  $n > n_0$ :

$$\Pr[A(Y, (U, V)) = (Y, \phi, \psi)(U, V)] < \frac{1}{n^c} .$$

*Proof.* The proof is by contradiction. Assume  $A, c > 0, \phi$ , and  $\psi$ , as above shows the proposition false for indices  $n$  in some infinite index set  $\mathcal{N}$ . Define a function  $f_s$  for each  $s \in \mathbb{Z}_q$  by  $f_s(g^{m\epsilon}) = g^{\psi_s(m\epsilon)}$ .

We describe a probabilistic circuit family  $A'$  that uses  $A$  to compute the knowledge function  $f_s$  with notable probability. This breaks the semantic security of the El Gamal cryptosystem, if  $p_n(f_s, M)$  is negligible. Given input  $(y, (u, v))$ , where  $(u, v) = (g^r, y^r m) \in G \times G$ ,  $A'$  does the following:

1. It randomly chooses  $s \in \mathbb{Z}_q$ .
2. It uses  $A$  to compute  $(u_0, v_0) = A(y, (u, v))$  and  $(u_1, v_1) = A(y, (u, v g^s))$
3. It returns  $\frac{v_1}{v_0}$ .

Let  $S = \{S_n\}$  be a u.i.d. random variable over  $\mathbb{Z}_q$ , and let  $H_0$  denote the event that  $A(Y, (U, V)) = (Y, \phi, \psi)(U, V)$ , and  $H_1$  denote the event that  $A(Y, (U, V g^S)) = (Y, \phi, \psi)(U, V g^S)$ .

If the events  $H_0$  and  $H_1$  take place we have  $\frac{v_1}{v_0} = f_S(M)$  by definition of the algorithm.

We see that  $((U, V)|R = r)$  and  $((U, V g^S)|R = r)$  are independent variables. Since  $R$  is u.i.d. we have:

$$\begin{aligned} \Pr[H_0 \wedge H_1] &= \sum_{r \in \mathbb{Z}_q} \Pr[R = r] \Pr[H_0 \wedge H_1 | R = r] \\ &= \sum_{r \in \mathbb{Z}_q} \Pr[R = r] \Pr[H_0 | R = r]^2 \\ &\geq \left( \sum_{r \in \mathbb{Z}_q} \Pr[R = r] \Pr[H_0 | R = r] \right)^2 \\ &= \Pr[H_0]^2 \geq \frac{1}{n^{2c}} \end{aligned}$$

where the inequality is implied by the convexity of the function  $h(x) = x^2$  and Jensen's Inequality.

We are only interested in outcomes  $s$  of  $S$  such that  $p_n(\psi_s, M) = p_n(f_s, M)$  is negligible (in particular  $s \neq 0$ ). Let  $W$  denote the event that  $S$  has this property. By assumption the probability of  $\overline{W}$  is negligible and we have:

$$\Pr[W \wedge A'(Y, (U, V)) = f_S(M)] \geq \frac{1}{2n^{2c}} .$$

The inequality implies that there exists for each  $n \in \mathcal{N}$  an outcome  $s_n$  of  $S_n$  such that the inequality still holds. Let  $A'' = \{A''_n\}$  be the circuit family that is identical to  $A'$  except that  $A''_n$  uses this fixed  $s_n$  instead of choosing it randomly. We set  $s = \{s_n\}$  and  $f_s = \{f_{s_n}\}$ , and conclude that  $A''$  has the property:

$$\Pr[A''(Y, (U, V)) = f_s(M)] \geq \frac{1}{2n^{2c}} ,$$

for  $n \in \mathcal{N}$ . Semantic security of the El Gamal cryptosystem implies that  $\forall c' > 0$ ,  $\exists n_0$  such that for  $n > n_0$  holds:

$$\Pr[A''(Y, (U, V)) = f_s(M)] < p_n(f_s, M) + \frac{1}{n^{c'}} .$$

Since  $f_s$  was constructed such that  $p_n(f_s, M)$  is negligible we have reached a contradiction.  $\square$

The proposition can be slightly generalized by considering distributions of the messages that are only almost uniform on its support when the support is sufficiently large. To keep this note simple we omit this analysis.

We proceed by defining a special subclass of the strongly non-linear functions that is particularly simple, and may be important in applications.

**Definition 4.** Let  $\psi = \{\psi_n\}$  be a family of functions,  $\psi_n : \mathbb{Z}_{q_n} \rightarrow \mathbb{Z}_{q_n}$ . We say that  $\psi$  has low degree if  $\forall c > 0$ ,  $\exists n_0$  such that for  $n > n_0$  it holds that:

$$\frac{\deg \psi_n}{q_n} < \frac{1}{n^c} .$$

A simple example of a family  $\psi = \{\psi_n\}$  that satisfies the above definition is where  $\psi_n(x) = p(x)$  for some fixed polynomial  $p(x)$  for all  $n$ .

We have the following corollary almost immediately from the proposition.

**Corollary 1.** Let  $G, X, Y, M, (U, V), \phi$  and  $\psi$  be as in Hypothesis 1, let  $M$  be u.i.d. in  $G$ , and assume that  $\psi$  has low degree and that  $\deg \psi_n \leq 1$  for at most finitely many  $n$ .

Then  $\forall A \in \text{PC}$ ,  $\forall c > 0$ ,  $\exists n_0 > 0$ , such that for  $n > n_0$ :

$$\Pr[A(Y, (U, V)) = (Y, \phi, \psi)(U, V)] < \frac{1}{n^c} .$$

*Proof.* It suffices to show that if  $\psi$  has low degree and  $\deg \psi_n \leq 1$  for finitely many  $n$  then  $\psi$  is strongly non-linear. For  $s \neq 0$  and large enough  $n$  we have  $\deg \psi_s > 0$  and  $\deg \psi_s = \deg \psi - 1$ . This implies that when  $s \neq 0$  we have  $p_n(\psi_s, M) = \frac{\max |\psi_s^{-1}(v)|}{q_n} \leq \frac{\max |\psi^{-1}(v)|}{q_n} \leq \frac{\deg \psi}{q_n}$ , which is negligible since  $\psi$  has low degree.  $\square$



### 3 Conclusion

It seems impossible to prove anything useful about general malleability of the El Gamal cryptosystem as discussed in Section 1.3. Instead we have formalized what we consider a reasonably restricted problem.

Under these restrictions we have exhibited a class of transformations that are not feasible to compute, when the message distribution is uniform. This may be of practical value when arguing about the security of certain protocols based on El Gamal. We have also given examples that indicate that the best possible results are not as strong as one may think.

It is an open problem to characterize further classes of transformations. A natural generalization is to consider lists of cryptotexts and consider the difficulty of computing transformations on such lists. This and other generalizations are relevant for mix-nets, and mix-net based voting schemes, where robustness is based on repetition and the impossibility of general transformations.

Another interesting line of research is to investigate the malleability properties of El Gamal in concrete groups, e.g. the multiplicative group of integers modulo a prime, or an elliptic curve group.

### 4 Acknowledgement

We had discussions with Gunnar Sjödin and Torsten Ekedahl. Johan Håstad gave advice on how to simplify this note.

### References

- [1] Dan Boneh, *The Decision Diffie-Hellman Problem*, Proceedings of the Third Algorithmic Number Theory Symposium, Lecture Notes in Computer Science 1423, pp. 48–63, Springer-Verlag, 1998. 177
- [2] Taher El Gamal, *A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms*, IEEE Transactions on Information Theory 31, pp. 469–472, 1985. 176
- [3] Shafi Goldwasser, Silvio Micali, *Probabilistic Encryption*, Journal of Computer Science 28, pp. 270–299, 1984. 179
- [4] Yiannis Tsiounis, Moti Yung, *On the Security of El Gamal based Encryption*, International workshop on Public Key Cryptography, Lecture Notes in Computer Science 1431, pp. 117–134, Springer-Verlag, 1998. 177, 179